



University of North Florida  
**UNF Digital Commons**

---

UNF Graduate Theses and Dissertations

Student Scholarship

---

2020

## A Two-Stage Design for Comparing Binomial Treatments with a Standard

Cecelia K. Schmidt

University of North Florida, [cecelia.schmidt@unf.edu](mailto:cecelia.schmidt@unf.edu)

Follow this and additional works at: <https://digitalcommons.unf.edu/etd>



Part of the [Applied Statistics Commons](#), [Clinical Trials Commons](#), and the [Design of Experiments and Sample Surveys Commons](#)

---

### Suggested Citation

Schmidt, Cecelia K., "A Two-Stage Design for Comparing Binomial Treatments with a Standard" (2020).  
*UNF Graduate Theses and Dissertations*. 962.  
<https://digitalcommons.unf.edu/etd/962>

This Master's Thesis is brought to you for free and open access by the Student Scholarship at UNF Digital Commons. It has been accepted for inclusion in UNF Graduate Theses and Dissertations by an authorized administrator of UNF Digital Commons. For more information, please contact [Digital Projects](#).  
© 2020 All Rights Reserved



# A two-stage design for comparing binomial treatments with a standard

Cecelia Katherine Schmidt

A thesis submitted to the Department of Mathematics and Statistics  
in partial fulfillment of the requirements for the degree of

Master of Science in Mathematical Sciences

UNIVERSITY OF NORTH FLORIDA

COLLEGE OF ARTS AND SCIENCES

April, 2020

The thesis of Cecelia Katherine Schmidt is approved by:

---

Dr. Elena Buzaianu, Committee Chair

---

Date

---

Dr. Ping Sa, Committee Member

---

Date

---

Dr. Yisu Jia, Committee Member

---

Date

## Acknowledgements

I would like to thank Dr. Buzaianu for all of her help, expertise, and especially her patience throughout the writing process. I would also like to thank Dr. Sa and Dr. Jia for being on my thesis committee. Finally, I would like to thank my friends and family for listening to me ramble about this thesis.

# Contents

Acknowledgements	i
Abstract	iii
1 Introduction	1
2 Proposed design	2
2.1 Design procedure . . . . .	2
2.2 Power and size . . . . .	2
2.3 Monotonicity properties of the power . . . . .	4
2.4 Derivation of design parameters . . . . .	12
3 Numerical results, comparisons, and applications	15
3.1 Design parameters . . . . .	15
3.2 Comparison to other designs . . . . .	17
3.3 Example . . . . .	20
4 Conclusion	22
Appendix	24
Comparison tables . . . . .	24
Model code . . . . .	25
Simulation code . . . . .	28

## Abstract

We propose a method for comparing success rates of several populations among each other and against a desired standard success rate. This design is appropriate for a situation in which all experimental treatments have only two outcomes that can be considered “success” and “failure” respectively. The goal is to identify which treatment has the highest rate of success that is also higher than the desired standard.

The design combines elements of both hypothesis testing and statistical selection. At the first stage, if none of the samples have a number of successes above the appropriate standard for the design, the experiment is terminated before the second stage. If one or more of the samples do exceed the standard, we continue to the second stage and take another sample from the population with the highest success rate in stage one. If the second stage produces a test statistic that is greater than the cutoff value for the second stage, we conclude that its associated treatment group/population has the highest success rate, which is also higher than the standard.

Since this procedure is not a pure hypothesis testing procedure, power and size are redefined in order to account for the hybrid selection and hypothesis testing nature of the design. We determine the design parameters for any given size and power of the procedure. When multiple designs meet the requirements we will recommend the design that has the lowest expected sample size.

# 1 Introduction

This design provides a method for comparing success rates between several populations and a desired standard. It is used when treatment success rates are to be compared to a given standard rate, or when there is no identified control treatment to which the experimental treatments are to be compared. Inspired by the approach in [3], our design instead calculates exact probabilities using the binomial distribution, rather than by using a normal approximation. It also does not involve sampling from a control group, since we are instead comparing the success rates to a standard rate.

The null hypothesis is

$$H_0 : \theta_0 = \theta_1 = \dots = \theta_k$$

where  $\theta_0$  is the standard success rate and  $\theta_i$  is the success rate in the  $i^{th}$  population or treatment group, indicated by  $P_i$ . If none of the samples have a success rate above the appropriate standard for the design, the experiment is terminated before the second stage. Otherwise, we continue to the second stage, in which we only sample from the group with the highest success rate. This is to reduce the total sample size needed to perform the procedure.

The rest of the paper is organized as follows. In the second section, we describe the specifics of the design and prove that a certain configuration is the least favorable - that is, that the power is minimized for such  $\theta_i$  values. This least favorable configuration (LFC) is then used in our power and size calculations. Tables of design parameters are provided next, with a brief discussion of the derivation of the design parameters and several simulation studies. To provide further context for this research, we compare it to two other designs from Buzaiianu [2] and Thall et. al [3]. We conclude with a brief example of applying the design and directions for future study.

## 2 Proposed design

### 2.1 Design procedure

Let  $X_{ji}$  be the number of successes in the  $i^{th}$  stage from the  $j^{th}$  treatment, and let  $n_i$  be the number of experimental units in each treatment in the  $i^{th}$  stage.

Stage 1: Experimental units are randomly assigned to all  $k$  treatment groups, for a total of  $kn_1$  units in the first stage. Then if

$$X_1 = X_{\zeta 1} = \max_{1 \leq j \leq k} X_{j1} > y_1$$

we continue to the second stage. Otherwise, we terminate early and fail to reject  $H_0$ .

Stage 2: We randomly assign another  $n_2$  units to the population we selected in stage 1,  $P_\zeta$ . If

$$X_2 = X_{\zeta 1} + X_{\zeta 2} > y_2$$

we reject  $H_0$  and conclude that  $\theta_\zeta > \theta_0$ . Otherwise, we fail to reject  $H_0$ .

### 2.2 Power and size

Now our goal is to find values of each of the unknown design constants  $n_1$ ,  $n_2$ ,  $y_1$ , and  $y_2$ . Since this procedure involves hypothesis testing, we must also find values of these constants that allow us to achieve a certain power value. To do all of this, we must first define the power; since power is dependent on the success rates of the populations, we must establish some definitions for these success rates. An “acceptable” population has success rate  $\theta_i \geq \theta_0 + \delta_2$ , and an “unacceptable” population has  $\theta_i \leq \theta_0 + \delta_1$ , where  $\delta_2 > \delta_1$ . We can think of  $\delta_1$  as the maximum insignificant difference from  $\theta_0$  and  $\delta_2$  as the minimum significant difference from  $\theta_0$ .  $\delta_1$  and  $\delta_2$  are between 0 and 1, and  $\theta_0 + \delta_2 < 1$ . Note that  $\delta_1$  and  $\delta_2$  are predetermined



values chosen by the researcher performing the experiment. Knowing this, we make the following two assumptions.

1. There is only one “acceptable” treatment/population.
2. All other treatments are “unacceptable”.

For shorthand, we denote a general set of success rate values  $(\theta_1, \dots, \theta_k)$  that fulfills the two assumptions as  $\theta$ . That is, all configurations  $\theta$  will have the general form  $\theta_1 \leq \theta_2 \leq \dots \leq \theta_{k-1} \leq \theta_0 + \delta_1 < \theta_0 + \delta_2 \leq \theta_k$ , with no success rates between  $\theta_0 + \delta_1$  and  $\theta_0 + \delta_2$ . WLOG we represent the highest success rate with  $\theta_k$ , from population  $P_k$ , and arrange the  $\theta_i$  in increasing order. Under our assumptions, this is also the only “acceptable” population. We now define the power in a general form.

$$\begin{aligned}
\text{Power}(\theta) &= P(\text{selecting } P_k | \theta) \\
&= P(P_k \text{ selected at } 1^{st} \text{ stage} \cap P_k \text{ selected at } 2^{nd} \text{ stage} | \theta) \\
&= P(P_k \text{ selected at } 1^{st} \text{ stage})P(P_k \text{ selected at } 2^{nd} \text{ stage} | P_k \text{ selected at } 1^{st} \text{ stage}, \theta) \\
&= P(P_k \text{ selected at } 1^{st} \text{ stage, including ties} \cap P_k \text{ produces max number of} \\
&\quad \text{successes in } 1^{st} \text{ stage} \cap X_1 > y_1 | \theta)P(X_2 > y_2 | P_k \text{ selected at } 1^{st} \text{ stage}, \theta) \\
&= \sum_{x=0}^{n_1} P(X_{k1} = x | \theta_k)I(x > y_1)P(X_{k2} + x > y_2 | \theta_k) \\
&\quad \cdot P(P_k \text{ has max number of successes} \cap P_k \text{ wins randomization if there} \\
&\quad \text{are ties} | X_{k1} = x, \theta)
\end{aligned}$$

This definition of power depends on the unknown set of success rates  $\theta$  and would thus be impossible to calculate for all possible configurations. Rather than attempting to do so, we are instead interested in finding a minimum value of power such that all configurations that fulfill our two assumptions will be guaranteed to have at least that much power. The values  $\theta$  associated with this least value of power are known as the least favorable configuration, or the LFC.

Now that we have defined power, we will also define size, or the Type I error rate. The

size,  $\alpha$ , is the probability of rejecting the null hypothesis when it is true. In our case, this translates into the probability of selecting any of the  $k$  populations under the null hypothesis configuration, that is, when  $\theta_0 = \theta_1 = \dots = \theta_k$ . Since the above power formula is  $P(\text{selecting } P_k | \theta)$ , the probability of selecting one of the populations, we can use this expression to provide a formula for  $\alpha$ .

### 2.3 Monotonicity properties of the power

To find the LFC, we will examine the behavior of the power relative to each of the success rates included in  $\theta$ . If we can find that the power is increasing or decreasing with respect to certain  $\theta_i$ ,  $i = 1, \dots, k$ , we may be able to find a minimal point. In the following proofs, we denote the last factor in the final power summation above in the shorter form  $P(W|x, \theta)$ . Additionally, we refer to the binomial PDF as  $b(x, n, p)$  and the binomial CDF as  $B(x, n, p)$ .

*Theorem 1. Power( $\theta$ ) is decreasing with respect to each  $\theta_i, i = 1, \dots, k - 1$ , when all other  $\theta_j, j \neq i$  are held constant.*

*Proof.* Since  $P(W|x, \theta)$  is the only part of the power formula that includes  $\theta_1, \dots, \theta_{k-1}$ , we consider this part alone. We should note that we are able to take the derivative with respect to  $\theta_i$  since these functions are continuous in  $\theta_i$  on the interval  $(0, 1)$ . We will show WLOG that the power is decreasing with respect to  $\theta_1$ . This probability can be split into the two following cases:

1. If  $x = 0$ : This is the case that all samples yield an  $X_{i1}$  of 0, so  $P(W|x, \theta)$  is the probability that  $P_k$  wins such a tie.

$$P(W|x, \theta) = \frac{1}{k} \prod_{i=1}^{k-1} b(0, n_1, \theta_i) = \frac{1}{k} \prod_{i=1}^{k-1} \binom{n_1}{0} \theta_i^0 (1 - \theta_i)^{n_1} = \frac{1}{k} \prod_{i=1}^{k-1} (1 - \theta_i)^{n_1}$$

We take the derivative with respect to  $\theta_1$ .

$$\frac{\partial}{\partial \theta_1} P(W|x, \theta) = \frac{\partial}{\partial \theta_1} \left[ \frac{1}{k} \prod_{i=1}^{k-1} (1 - \theta_i)^{n_1} \right]$$

$$\begin{aligned}
&= \left( \frac{1}{k} \prod_{i=2}^{k-1} (1 - \theta_i)^{n_1} \right) \frac{\partial}{\partial \theta_1} [(1 - \theta_1)^{n_1}] \\
&= \left( \frac{1}{k} \prod_{i=2}^{k-1} (1 - \theta_i)^{n_1} \right) (n_1 (1 - \theta_1)^{n_1-1} (-1)) < 0
\end{aligned}$$

Thus, for  $x = 0$ ,  $P(W|x, \theta)$  is decreasing in  $\theta_1$ , and therefore also decreasing with respect to any  $\theta_i$ , for  $1 \leq i \leq k-1$ . We must now consider the remaining case.

2. If  $x > 0$ : This situation has to include all ways in which  $P_k$ 's sample can either have the highest  $X_{i1}$  value or win a tie between any number of other populations.

$$\begin{aligned}
P(W|x, \theta) &= \prod_{j=1}^{k-1} B(x-1, n_1, \theta_j) \\
&\quad + \frac{1}{2} \sum_{i=1}^{\binom{k-1}{1}} \prod_{j \in A_i^1} b(x, n_1, \theta_j) \prod_{\substack{l \notin A_i^1 \\ 1 \leq l \leq k-1}} B(x-1, n_1, \theta_l) \\
&\quad + \dots + \frac{1}{k-1} \sum_{i=1}^{\binom{k-1}{k-2}} \prod_{j \in A_i^{k-2}} b(x, n_1, \theta_j) \prod_{\substack{l \notin A_i^{k-2} \\ 1 \leq l \leq k-1}} B(x-1, n_1, \theta_l) \\
&\quad + \frac{1}{k} \prod_{j=1}^{k-1} b(x, n_1, \theta_j)
\end{aligned}$$

where  $A^t = \{\text{all unordered } t\text{-tuples from } \{1, 2, \dots, k-1\}\} = \{A_1^t, A_2^t, \dots, A_{\binom{k-1}{t}}^t\}$ , and we define  $A^0 = \emptyset$ ,  $A^{k-1} = \{(1, 2, \dots, k-1)\}$ .  $A^0$  and  $A^{k-1}$  are implicitly used in the first and last terms of the summation, which we can see if we write the probability in the general form

$$P(W|x, \theta) = \sum_{t=0}^{k-1} \frac{1}{t+1} \sum_{i=1}^{\binom{k-1}{t}} \prod_{j \in A_i^t} b(x, n_1, \theta_j) \prod_{\substack{l \notin A_i^t \\ 1 \leq l \leq k-1}} B(x-1, n_1, \theta_l)$$

Note that this summation has  $\sum_{t=0}^{k-1} \binom{k-1}{t} = 2^{k-1}$  terms, so it always has an even number of terms. Also, each individual term has a “paired” term, which has every factor in common with another term except for some form of  $\frac{1}{t+1}$  and the factor  $b(x, n_1, \theta_1)$ , which is replaced in the “paired” term by  $B(x-1, n_1, \theta_1)$ . These two terms are the same in every factor except for the PDF/CDF for  $\theta_1$  and the leading fraction. Such pairings, for any  $k$ , are what we

mean by a “paired” term. We also know that, since one of the terms in the pair has one more PDF than the other, that that term has a smaller factor of  $\frac{1}{t+1}$  (that is, a larger value of  $t+1$ ). This is because the term with one more PDF represents one more sample that produced a tie with  $P_k$ , so we must now randomly break the tie between more populations. Of course, we want to ensure that this pairing process will work in all of our possible situations.

It is straightforward to show that this happens when there are only  $k = 2$  populations. As a reminder, we assume that we are pairing terms that only differ in the PDF/CDF for  $\theta_1$ . Either  $P_1$  ties with  $P_2$ , or it does not. For this simpler case, our summation only has two terms, which are immediately paired with each other:

$$B(x-1, n_1, \theta_1) + \frac{1}{2}b(x, n_1, \theta_1)$$

We extend this to an inductive argument on the number of populations, by thinking of each individual term in our  $P(W|x, \theta)$  summation as representing one of the possible subsets of the set  $[k-1] = \{1, 2, \dots, k-1\}$ , following a similar strategy to a proof in [1, p. 27]. This relationship can be seen in our definition of  $A^t$ , since the number and location of the ties/PDFs is determined by which subset of  $[k-1]$  the  $A_i^t$  matches. In the  $k = 2$  case shown above, our set  $[k-1]$  is just  $\{1\}$ . The first term in the summation is associated with the subset  $\emptyset$ , because there are no ties, and the second term is associated with the subset  $\{1\}$ , because there is one tie.

Assume that this “pairing” is possible for  $k$  populations, or all  $2^{k-1}$  subsets of  $[k-1]$ . There are twice as many subsets of  $[k]$  when we go up to  $k+1$  populations, which are either subsets of  $[k-1]$  (all  $2^{k-1}$  of them), or subsets of  $[k-1]$  with  $k$  added, which is the other  $2^{k-1}$  subsets, adding up to all  $2^k$  subsets. Since we already know that the subsets of  $[k-1]$  can be paired with each other, the subsets of  $[k]$  are paired in the same way. In the case of  $k+1 = 3$  populations, it works as follows:  $\emptyset$  is paired with  $\{1\}$ , and  $\{2\}$  is paired with  $\{1, 2\}$ , to produce

$$B(x-1, n_1, \theta_1)B(x-1, n_2, \theta_2) + \frac{1}{2}b(x, n_1, \theta_1)B(x-1, n_1, \theta_2) \text{ (paired)}$$

$$+ \frac{1}{2}B(x-1, n_1, \theta_1)b(x, n_1, \theta_2) + \frac{1}{3}b(x, n_1, \theta_1)b(x, n_1, \theta_2) \quad (\text{paired})$$

Now that we know how these pairs of terms are related to each other, we can factor every such pairing into the form below, where the contents of the product are different depending on which pair of terms we are factoring. We also want to manipulate these two terms into a form that is easier to show as being decreasing with respect to  $\theta_1$ .

$$\begin{aligned} & \left( \frac{1}{t+2}b(x, n_1, \theta_1) + \frac{1}{t+1}B(x-1, n_1, \theta_1) \right) \prod[\dots] \\ &= \left( (t+1)b(x, n_1, \theta_1) + (t+2)B(x-1, n_1, \theta_1) \right) \frac{1}{(t+1)(t+2)} \prod[\dots] \\ &= \left( (t+1)B(x, n_1, \theta_1) + B(x-1, n_1, \theta_1) \right) \prod[\dots] \end{aligned}$$

Since this can be done to all pairs of terms, we can rewrite  $P(W|x, \theta)$  into a summation of pairs of binomial CDF terms for  $\theta_1$ , where each pair is multiplied by a product with no factors of  $\theta_1$ . We now show that the binomial CDF  $B(x, n_1, \theta_1)$  is decreasing with respect to  $\theta_1$ . The binomial CDF has the form

$$B(x, n_1, \theta_1) = \sum_{i=0}^x \binom{n_1}{i} \theta_1^i (1 - \theta_1)^{n_1-i}$$

We take the derivative with respect to  $\theta_1$ .

$$\begin{aligned} & \frac{\partial}{\partial \theta_1} \left[ \sum_{i=0}^x \binom{n_1}{i} \theta_1^i (1 - \theta_1)^{n_1-i} \right] \\ &= \sum_{i=0}^x \binom{n_1}{i} (i\theta_1^{i-1}(1 - \theta_1)^{n_1-i} - (n_1 - i)\theta_1^i(1 - \theta_1)^{n_1-i-1}) \end{aligned}$$

Now we select two arbitrary consecutive terms in the summation. Note that when we write these two terms in this manner, the index  $i$  cannot be greater than  $n_1 - 1$ .

$$\begin{aligned} & \binom{n_1}{i} (i\theta_1^{i-1}(1 - \theta_1)^{n_1-i} - (n_1 - i)\theta_1^i(1 - \theta_1)^{n_1-i-1}) \\ &+ \binom{n_1}{i+1} ((i+1)\theta_1^i(1 - \theta_1)^{n_1-i-1} - (n_1 - i - 1)\theta_1^{i+1}(1 - \theta_1)^{n_1-i-2}) \\ &= \frac{n_1!}{i!(n_1 - i)!} i\theta_1^{i-1}(1 - \theta_1)^{n_1-i} - \frac{n_1!}{i!(n_1 - i)!} (n_1 - i)\theta_1^i(1 - \theta_1)^{n_1-i-1} \\ &+ \frac{n_1!}{(i+1)!(n_1 - i - 1)!} (i+1)\theta_1^i(1 - \theta_1)^{n_1-i-1} \end{aligned}$$

$$\begin{aligned}
& - \frac{n_1!}{(i+1)!(n_1-i-1)!} (n_1-i-1) \theta_1^{i+1} (1-\theta_1)^{n_1-i-2} \\
& = \frac{n_1!}{i!(n_1-i)!} i \theta_1^{i-1} (1-\theta_1)^{n_1-i} - \frac{n_1!}{(i+1)!(n_1-i-1)!} (n_1-i-1) \theta_1^{i+1} (1-\theta_1)^{n_1-i-2}
\end{aligned}$$

Knowing that this is a telescoping sum, we are left with only parts of the first and last terms.

$$\begin{aligned}
\frac{\partial}{\partial \theta_1} B(x, n_1, \theta_1) & = \binom{n_1}{0} (0) \theta_1^{0-1} (1-\theta_1)^{n_1-0} - \binom{n_1}{x} (n_1-x) \theta_1^x (1-\theta_1)^{n_1-x-1} \\
& = - \binom{n_1}{x} (n_1-x) \theta_1^x (1-\theta_1)^{n_1-x-1} < 0
\end{aligned}$$

Therefore, for  $x > 0$ ,  $P(W|x, \theta)$  is decreasing in terms of  $\theta_1$ , and this is extended to all  $\theta_1, \theta_2, \dots, \theta_{k-1}$ . ■

**Theorem 2.** *Power( $\theta$ ) is increasing with respect to  $\theta_k$  when all other  $\theta_i, i = 1, \dots, k-1$  are held constant.*

*Proof.* The only part of the power summation that deals with  $\theta_k$  is  $P(X_{k1} = x | \theta_k) P(X_{k1} + x > y_2 | \theta_k)$ , but unlike with the decreasing case, we will use the whole power formula.

$$\begin{aligned}
\frac{\partial}{\partial \theta_k} [\text{Power}(\theta)] & = \frac{\partial}{\partial \theta_k} \left[ \sum_{x=0}^{n_1} P(X_{k1} = x | \theta_k) I(x > y_1) P(X_{k2} + x > y_2 | \theta_k) P(W|x, \theta) \right] \\
& = \sum_{x=0}^{n_1} P(X_{k1} = x | \theta_k) I(x > y_1) \frac{\partial}{\partial \theta_k} [P(X_{k2} + x > y_2 | \theta_k)] P(W|x, \theta) \quad (1) \\
& \quad + \sum_{x=0}^{n_1} \frac{\partial}{\partial \theta_k} [P(X_{k1} = x | \theta_k)] I(x > y_1) P(X_{k2} + x > y_2 | \theta_k) P(W|x, \theta) \quad (2)
\end{aligned}$$

First we examine (1). All of the factors outside the derivative must be positive, assuming the indicator is nonzero, so we look at the derivative part. We are left with a summation that we already showed to be telescoping on the previous page. Note that if  $y_2 - x \leq 0$ , this derivative will just be 0, so we are focusing on the case that this quantity is strictly positive.

$$\begin{aligned}
\frac{\partial}{\partial \theta_k} [P(X_{k2} + x > y_2 | \theta_k)] & = \frac{\partial}{\partial \theta_k} [P(X_{k2} > y_2 - x | \theta_k)] \\
& = \frac{\partial}{\partial \theta_k} \left[ \sum_{i=y_2-x+1}^{n_2} \binom{n_2}{i} \theta_k^i (1-\theta_k)^{n_2-i} \right] \\
& = \sum_{i=y_2-x+1}^{n_2} \binom{n_2}{i} [i \theta_k^{i-1} (1-\theta_k)^{n_2-i} - (n_2-i) \theta_k^i (1-\theta_k)^{n_2-i-1}]
\end{aligned}$$

$$\begin{aligned}
&= \binom{n_2}{y_2 - x + 1} (y_2 - x + 1) \theta_k^{y_2 - x} (1 - \theta_k)^{n_2 - y_2 + x - 1} \\
&\quad - \binom{n_2}{n_2} (n_2 - n_2) \theta_k^{n_2} (1 - \theta_k)^{n_2 - n_2} \\
&= \binom{n_2}{y_2 - x + 1} (y_2 - x + 1) \theta_k^{y_2 - x} (1 - \theta_k)^{n_2 - y_2 + x - 1} > 0
\end{aligned}$$

Therefore, (1) is positive. Now we look at (2), defining a function  $f(x) = I(x > y_1)P(X_{k2} + x > y_2 | \theta_k)P(W|x, \theta)$  for shorthand.

$$\begin{aligned}
&\sum_{x=0}^{n_1} \frac{\partial}{\partial \theta_k} [P(X_{k1} = x | \theta_k)] f(x) \\
&= \sum_{x=0}^{n_1} f(x) \frac{\partial}{\partial \theta_k} \left[ \binom{n_1}{x} \theta_k^x (1 - \theta_k)^{n_1 - x} \right] \\
&= \sum_{x=0}^{n_1} f(x) \binom{n_1}{x} (x \theta_k^{x-1} (1 - \theta_k)^{n_1 - x} - (n_1 - x) \theta_k^x (1 - \theta_k)^{n_1 - x - 1}) \\
&= f(0) \binom{n_1}{0} ((0) \theta_k^{0-1} (1 - \theta_k)^{n_1} - n_1 \theta_k^0 (1 - \theta_k)^{n_1 - 1}) \\
&\quad + f(1) \binom{n_1}{1} ((1) \theta_k^0 (1 - \theta_k)^{n_1 - 1} - (n_1 - 1) \theta_k^1 (1 - \theta_k)^{n_1 - 2}) \\
&\quad + \dots + \\
&\quad + f(n_1) \binom{n_1}{n_1} ((n_1) \theta_k^{n_1 - 1} (1 - \theta_k)^{n_1 - n_1} - (n_1 - n_1) \theta_k^{n_1} (1 - \theta_k)^{-1})
\end{aligned}$$

This is almost the same as our telescoping sum from earlier, but we now have an  $f(x)$  factor in every single term. We pair up the terms that cancelled previously and define a new function  $Q(x)$ . Note that in the next expression, the  $x$  will be at most  $n_1 - 1$ .

$$\begin{aligned}
&f(x) \binom{n_1}{x} [-(n_1 - x) \theta_k^x (1 - \theta_k)^{n_1 - x - 1}] + f(x + 1) \binom{n_1}{x + 1} [(x + 1) \theta_k^x (1 - \theta_k)^{n_1 - x - 1}] \\
&= f(x) \frac{n_1!}{x!(n_1 - x - 1)!} [-\theta_k^x (1 - \theta_k)^{n_1 - x - 1}] \\
&\quad + f(x + 1) \frac{n_1!}{x!(n_1 - x - 1)!} [\theta_k^x (1 - \theta_k)^{n_1 - x - 1}] \\
&= (f(x + 1) - f(x))Q(x)
\end{aligned}$$

We want to show that the above equation is positive, so we will show that  $f(x + 1) > f(x)$ .

Recall how we defined  $f(x+1)$ :

$$f(x+1) = I(x+1 > y_1)P(X_{k2} + (x+1) > y_2 \mid \theta_k)P(W|x+1, \theta)$$

First we examine the indicator function. Since it is defined by

$$I(x+1 > y_1) = \begin{cases} 1 & x+1 > y_1 \\ 0 & x+1 \leq y_1 \end{cases}$$

we notice that the values of  $I$  are zero up until  $x+1 > y_1$ , and then every value onward is one. In other words,  $I(x+1 > y_1) \geq I(x > y_1)$ .

The second factor is a binomial probability. It is straightforward to say that  $P(X_{k2} + (x+1) > y_2 \mid \theta_k) \geq P(X_{k2} + x > y_2 \mid \theta_k)$ .

We are left with  $P(W|x+1, \theta)$ . Now we have to split this into two cases:  $x = 0$  and  $x > 0$ , because the  $P(W|x, \theta)$  term has a different form for  $x = 0$ . For now, we look at the  $x > 0$  case and show that  $P(W|x+1, \theta) \geq P(W|x, \theta)$ . We recall (after some effort) that this probability is equal to

$$\begin{aligned} P(W|x+1, \theta) &= \prod_{j=1}^{k-1} B(x, n_1, \theta_j) \\ &\quad + \frac{1}{2} \sum_{i=1}^{\binom{k-1}{1}} \prod_{j \in A_i^1} b(x+1, n_1, \theta_j) \prod_{\substack{l \notin A_i^1 \\ 1 \leq l \leq k-1}} B(x, n_1, \theta_l) \\ &\quad + \dots + \frac{1}{k-1} \sum_{i=1}^{\binom{k-1}{k-2}} \prod_{j \in A_i^{k-2}} b(x+1, n_1, \theta_j) \prod_{\substack{l \notin A_i^{k-2} \\ 1 \leq l \leq k-1}} B(x, n_1, \theta_l) \\ &\quad + \frac{1}{k} \prod_{j=1}^{k-1} b(x, n_1, \theta_j) \end{aligned}$$

We look at the first term of this summation. It can be split as follows:

$$\begin{aligned} \prod_{j=1}^{k-1} B(x, n_1, \theta_j) &= \prod_{j=1}^{k-1} (b(x, n_1, \theta_j) + B(x-1, n_1, \theta_j)) \\ &= \prod_{j=1}^{k-1} B(x-1, n_1, \theta_j) \end{aligned}$$



$$\begin{aligned}
& + \sum_{i=1}^{\binom{k-1}{1}} \prod_{j \in A_i^1} b(x, n_1, \theta_j) \prod_{\substack{l \notin A_i^1 \\ 1 \leq l \leq k-1}} B(x-1, n_1, \theta_l) \\
& + \dots + \sum_{i=1}^{\binom{k-1}{k-2}} \prod_{j \in A_i^{k-2}} b(x, n_1, \theta_j) \prod_{\substack{l \notin A_i^{k-2} \\ 1 \leq l \leq k-1}} B(x-1, n_1, \theta_l) \\
& + \prod_{j=1}^{k-1} b(x, n_1, \theta_j) \\
& > \prod_{j=1}^{k-1} B(x-1, n_1, \theta_j) \\
& + \frac{1}{2} \sum_{i=1}^{\binom{k-1}{1}} \prod_{j \in A_i^1} b(x, n_1, \theta_j) \prod_{\substack{l \notin A_i^1 \\ 1 \leq l \leq k-1}} B(x-1, n_1, \theta_l) \\
& + \dots + \frac{1}{k-1} \sum_{i=1}^{\binom{k-1}{k-2}} \prod_{j \in A_i^{k-2}} b(x, n_1, \theta_j) \prod_{\substack{l \notin A_i^{k-2} \\ 1 \leq l \leq k-1}} B(x-1, n_1, \theta_l) \\
& + \frac{1}{k} \prod_{j=1}^{k-1} b(x, n_1, \theta_j) = P(W|x, \theta)
\end{aligned}$$

Therefore,  $P(W|x+1, \theta) > \prod_{j=1}^{k-1} B(x, n_1, \theta_j) > P(W|x, \theta)$ , and we have that  $f(x+1) > f(x) \forall x = 1, 2, 3, \dots$ . We now address  $f(0)$  and  $f(1)$ . Specifically, we compare  $P(W|0, \theta)$  and  $P(W|1, \theta)$ , recalling how we just split  $P(W|x+1, \theta)$ .

$$\begin{aligned}
P(W|0, \theta) &= \frac{1}{k} \prod_{i=1}^{k-1} b(0, n_1, \theta_i) \\
P(W|1, \theta) &= \prod_{j=1}^{k-1} B(0, n_1, \theta_j) + \dots = \prod_{j=1}^{k-1} b(0, n_1, \theta_j) + \dots
\end{aligned}$$

So we have also that  $P(W|1, \theta) > P(W|0, \theta)$ , and can now say  $f(x+1) > f(x) \forall x = 0, 1, 2, \dots$ . Therefore, we have shown that the power summation is increasing with respect to  $\theta_k$ , when all other  $\theta_i$  are held constant. ■

Remembering our original two assumptions and definitions at the beginning of section 2.1, we can now derive the LFC. It is most important that we know that  $\theta_0 + \delta_1$  is the

upper bound on the “unacceptable” success rates, while  $\theta_0 + \delta_2$  is the lower bound on the “acceptable” success rates. Since we only have the one “acceptable” population, and the others are all “unacceptable”, we have the following theorem.

*Theorem 3. Out of all possible configurations  $\theta_1 \leq \theta_2 \leq \dots \leq \theta_{k-1} \leq \theta_0 + \delta_1 < \theta_0 + \delta_2 \leq \theta_k$ , power of the design is minimized under the configuration  $\theta_1 = \theta_2 = \dots = \theta_{k-1} = \theta_0 + \delta_1 < \theta_0 + \delta_2 = \theta_k$ .*

*Proof.* Starting with a general  $\theta$  configuration,

$$\theta_1 \leq \theta_2 \leq \dots \leq \theta_{k-1} \leq \theta_0 + \delta_1 < \theta_0 + \delta_2 \leq \theta_k$$

we know that increasing the “unacceptable” population success rates will decrease the power, and decreasing the “acceptable” population success rates will also decrease the power. Therefore, we perform these actions one at a time:

$$\begin{aligned} & \text{Power}(\theta_1 \leq \theta_2 \leq \dots \leq \theta_{k-1} \leq \theta_0 + \delta_1 < \theta_0 + \delta_2 \leq \theta_k) \\ & \geq \text{Power}(\theta_1 \leq \theta_2 \leq \dots \leq \theta_{k-1} \leq \theta_0 + \delta_1 < \theta_0 + \delta_2 = \theta_k) \\ & \geq \text{Power}(\theta_1 \leq \theta_2 \leq \dots \leq \theta_{k-1} = \theta_0 + \delta_1 < \theta_0 + \delta_2 = \theta_k) \\ & \dots \\ & \geq \text{Power}(\theta_1 = \theta_2 = \dots = \theta_{k-1} = \theta_0 + \delta_1 < \theta_0 + \delta_2 = \theta_k) \end{aligned} \quad \blacksquare$$

So, our LFC is

$$\theta_1 = \theta_2 = \dots = \theta_{k-1} = \theta_0 + \delta_1 < \theta_0 + \delta_2 = \theta_k$$

Note that the LFC does not depend on the unknown values of  $\theta$ , as long as they fulfill the specified assumptions.

## 2.4 Derivation of design parameters

Now that we have derived the LFC for our procedure, we can write the power formula in terms of our LFC, with specific parameter values. Since this is a lower bound on power, we

know that the procedure will be at least this powerful when our two assumptions are met.

$$\text{Power}(\theta) = \sum_{x=0}^{n_1} P(X_{k1} = x | \theta_k) I(X_{k1} > y_1) P(X_{k2} + x > y_2 | \theta_k) P(W|x, \theta)$$

We can now explicitly define each individual factor in the terms of the summation under the LFC, skipping over the indicator function since it is already clearly defined. We substitute values of  $\theta_i$  where appropriate, as defined under the LFC. The values below are used to find the minimal power of a design for given design parameters  $n_1$ ,  $n_2$ ,  $y_1$ , and  $y_2$ .

$$P(X_{k1} = x | \theta_k) = P(X_{k1} = x | \theta_0 + \delta_2) = b(x, n_1, \theta_0 + \delta_2)$$

$$P(X_{k2} + x > y_2 | \theta_k) = P(X_{k2} > y_2 - x | \theta_0 + \delta_2) = 1 - B(y_2 - x, n_2, \theta_0 + \delta_2)$$

$$\begin{aligned} P(W|x, \theta) &= \frac{1}{k} (b(0, n_1, \theta_0 + \delta_1))^{k-1} \text{ if } x = 0 \\ &= \sum_{i=0}^{k-1} \frac{1}{i+1} \binom{k-1}{i} (b(x, n_1, \theta_0 + \delta_1))^i (B(x-1, n_1, \theta_0 + \delta_1))^{k-1-i} \text{ if } x > 0 \end{aligned}$$

Using the power expression, we will now provide a formula for the size. Recall that size is the probability of selecting any of the populations under the null hypothesis. Due to symmetry,

$$\alpha = k \cdot P(\text{selecting } P_k | H_0)$$

which can be obtained by evaluating the above power expression when  $\delta_1 = \delta_2 = 0$ .

When we derive the design, it is possible that we will find multiple designs that fulfill the size and power requirements, especially as we increase the overall sample size. In this case, we want to select the design with the lowest expected sample size, since in a real-world situation a very large sample can mean a very large cost to perform our study. We define expected sample size in the same manner as [3].

$$E[N] = \frac{1}{2} [E[N|H_0] + E[N|LFC]]$$

We must define these two conditional expectation formulas. To do so, we generalize the

expected sample size under any configuration:

$$\begin{aligned}
E[N|\theta] &= kn_1 + n_2 P(X_1 > y_1 | \theta) \\
P(X_1 > y_1 | \theta) &= P(\max(X_{11}, \dots, X_{k1}) > y_1 | \theta) \\
&= 1 - P(\max(X_{11}, \dots, X_{k1}) \leq y_1 | \theta) \\
&= 1 - P(X_{11} \leq y_1 \cap \dots \cap X_{k1} \leq y_1 | \theta) \\
&= 1 - P(X_{11} \leq y_1) \dots P(X_{k1} \leq y_1 | \theta) \\
&= 1 - \prod_{i=1}^k P(X_{i1} \leq y_1 | \theta_i)
\end{aligned}$$

Now all we do is plug in the appropriate configurations ( $H_0$  and LFC), and get the final formula for expected sample size.

$$\begin{aligned}
E[N] &= \frac{1}{2} \left[ kn_1 + n_2 \left( 1 - [B(y_1, n_1, \theta_0)]^k \right) + kn_1 + n_1 \left( 1 - \prod_{i=1}^k B(y_1, n_1, \theta_i) \right) \right] \\
&= \frac{1}{2} \left[ 2kn_1 + n_2 (2 - [B(y_1, n_1, \theta_0)]^k - [B(y_1, n_1, \theta_0 + \delta_1)]^{k-1} B(y_1, n_1, \theta_0 + \delta_2)) \right]
\end{aligned}$$

While this next value is not used to select a design, it is useful for us to know the probability of termination before the second stage when the null hypothesis is true. We would like this to be somewhat high, that way we can be relatively certain that we do not waste resources testing for a non-existent difference.

$$\begin{aligned}
\tau_0 &= P(\text{stopping before the second stage}) \\
&= P(X_1 \leq y_1 | H_0) \\
&= P(\max(X_{11}, \dots, X_{k1}) \leq y_1 | \theta_1 = \dots = \theta_k = \theta_0) \\
&= P(X_{11} \leq y_1 \cap \dots \cap X_{k1} \leq y_1 | \theta_1 = \dots = \theta_k = \theta_0) \\
&= \prod_{i=1}^k P(X_{i1} \leq y_1 | \theta_0) \\
&= [B(y_1, n_1, \theta_0)]^k
\end{aligned}$$

We take all of these formulas and use them to compute the design parameters.

### 3 Numerical results, comparisons, and applications

#### 3.1 Design parameters

The design parameters were found using an R algorithm contained in the appendix. Values of  $k$ ,  $\theta_0$ , and power were entered for each design, and then the code ran through several positive integer values for each constant  $n_1$ ,  $n_2$ ,  $y_1$ , and  $y_2$ . Upper and lower bounds for possible values of these constants were determined by the values found in the design in [3], values found for designs with lower power, and “reasonable” considerations. By “reasonable” we mean that if there are several combinations of parameter values that fulfill the design constraints, significantly larger sample sizes will be unlikely to have a smaller expected sample size.

We propose values of  $n_1$ ,  $n_2$ ,  $y_1$ , and  $y_2$  for several possible combinations of values of the user-determined design parameters. For the number of treatments  $k$ : 2, 3, and 4. Standard success rate values  $\theta_0$  were 0.5, 0.6, and 0.7. Minimum power values were 0.7, 0.8, and 0.9. Size  $\alpha$  was set to 0.05 for all designs, as was  $\delta_1 = 0.05$  and  $\delta_2 = 0.2$ . Only potential designs with power at least equal to the desired value and size at most equal to the desired size value were considered. We present the ‘optimal’ set of design parameters; in our case, as we briefly discussed in the last section, we define ‘optimal’ as having the lowest expected sample size. Accordingly, the design with the lowest value of  $E[N]$  is the one displayed in the table on the next page. The last column contains the probabilities of early termination before the second stage.

Having derived these values, we checked a few of the values using simulation of the behavior of the design under the LFC. We chose to run simulation studies on the  $k = 2$  designs and recorded the results in table 2. The simulated power was calculated by simulating

the design and recording whether or not the “acceptable” population was selected at the end of stage 2. Each design was simulated 100,000 times.

Since we ran a random simulation, case-by-case results will differ slightly each time the program is run. The largest absolute difference between the theoretical power value and the simulated power value is 0.0024, for the first design. These values are sufficiently close for us to be satisfied with our power estimation. The same is true for our size values, and the algorithm used to estimate power (found in the appendix) can be quickly modified to output estimations for size if desired.

*Table 1:* Designs with minimal  $E[N]$ , for  $\alpha = 0.05$ ,  $\delta_1 = 0.05$ , and  $\delta_2 = 0.2$ .

$k$	$\theta_0$	Power	$n_1$	$n_2$	$y_1$	$y_2$	Power	Size	$E[N]$	$\tau_0$
2	0.5	0.7	18	32	10	31	0.7029	0.0472	57.3969	0.5771
		0.8	31	28	18	36	0.8002	0.0498	78.6136	0.7387
		0.9	45	33	24	47	0.9006	0.0482	114.2405	0.5245
	0.6	0.7	15	31	10	33	0.7010	0.0475	49.8542	0.6127
		0.8	24	26	15	36	0.8022	0.0489	67.9050	0.4517
		0.9	38	30	24	48	0.9006	0.0477	98.3475	0.5046
	0.7	0.7	13	17	10	25	0.7007	0.0471	36.8342	0.6360
		0.8	20	15	15	29	0.8021	0.0488	50.4501	0.5814
		0.9	29	19	22	39	0.9007	0.0481	70.6093	0.6594
3	0.5	0.7	23	42	13	40	0.7017	0.0465	99.3249	0.5073
		0.8	38	34	22	44	0.8037	0.0492	136.0958	0.6632
		0.9	58	37	33	57	0.9008	0.0479	198.1394	0.6849
	0.6	0.7	20	34	13	39	0.7041	0.0481	86.3269	0.4219
		0.8	30	36	20	47	0.8002	0.0483	115.4861	0.5589
		0.9	49	35	33	59	0.9000	0.0465	169.6553	0.6934
	0.7	0.7	17	20	13	31	0.7001	0.0415	65.5706	0.5083
		0.8	22	27	17	40	0.8021	0.0487	84.7439	0.5831
		0.9	36	22	28	47	0.9008	0.0495	122.1762	0.6994
4	0.5	0.7	28	47	16	46	0.7021	0.0476	147.2856	0.4690
		0.8	42	45	24	53	0.8012	0.0455	200.3412	0.5471
		0.9	67	34	38	61	0.9001	0.0477	291.3015	0.6252
	0.6	0.7	24	40	16	46	0.7006	0.0469	127.0001	0.4263
		0.8	36	40	24	54	0.8053	0.0498	173.9111	0.4931
		0.9	57	32	38	63	0.9003	0.0448	250.4226	0.5951
	0.7	0.7	19	27	15	38	0.7033	0.0446	94.7568	0.5646
		0.8	26	29	20	45	0.8017	0.0476	125.7009	0.4918
		0.9	41	27	32	55	0.9027	0.0472	181.8207	0.6729

Table 2: Simulation results for  $k = 2$

$k$	$\theta_0$	$n_1$	$n_2$	$y_1$	$y_2$	Exact power	Simulated power
2	0.5	18	32	10	31	0.7029	0.7052
		31	28	18	36	0.8002	0.8002
		45	33	24	47	0.9006	0.9016
	0.6	15	31	10	33	0.7010	0.6996
		24	26	15	36	0.8022	0.8020
		38	30	24	48	0.9006	0.9017
	0.7	13	17	10	25	0.7007	0.6998
		20	15	15	29	0.8021	0.8036
		29	19	22	39	0.9007	0.9002

### 3.2 Comparison to other designs

Because this design was inspired by Thall et. al's design in [3], we make a brief comparison between the two. Thall et. al's design should be applied when the researcher is interested in comparing treatments to a preexisting control treatment. That design also employs the normal approximation when deriving the design parameters. Our design should be used when the researcher is interested in comparing treatments to a known standard success rate, rather than sampling from a control. Also, our design uses exact binomial probabilities in all the calculations - so if a researcher has small sample size restrictions, they can still find the exact power of the design using a simpler form of the algorithm in the appendix. Of course, Thall et. al's design comes with a larger overall sample size, which is a consequence of sampling from the additional control population. Essentially, the choice depends on whether the researcher wants to use a control population or a standard.

We also compare this new design to a pair of designs proposed by Buzaiianu [2] for comparison to a standard success rate, both of which are pure selection procedures. The first is a single-stage, fixed sample size procedure; the other is a sequential design with curtailment that can be terminated as soon as a treatment has enough successes, or all treatments have enough failures. While the designs have slightly different formulations and goals than ours, they are comparable under certain parameter configurations. In order to compare the two, we must use five constants  $\delta_0^*, \delta_1^*, \delta_2^*, P_1^*, P_2^*$ , such that  $0 < \delta_1^*, \delta_2^* < \infty$ ,  $-\delta_1^* < \delta_0^* < \infty$ ,  $2^{-k} <$

$P_0^* < 1$ , and  $(1 - 2^{-k})k^{-1} < P_1^* < 1$  from [2]. We also define  $\Pi_0$  as the event of not selecting any of the populations, and  $\Pi_k$  as the event of selecting the population associated with  $\theta_k$ . Both procedures satisfy these inequalities:

$$P(\Pi_0|\theta_k \leq \theta_0 - \delta_0^*) \geq P_0^*$$

$$P(\Pi_k|\theta_k \geq \theta_{k-1} + \delta_1^*, \theta_k \geq \theta_0 + \delta_2^*) \geq P_1^*$$

These two designs each have two LFCs associated with them, one for each probability requirement listed above. The chosen design parameters are the smallest values fulfill both of these probability requirements.  $LFC_0$  is the configuration  $\theta_1 = \theta_2 = \dots = \theta_k = \theta_0 - \delta_0^*$ , and  $LFC_1$  is the configuration  $\theta_1 = \theta_2 = \dots = \theta_{k-1} = \theta_0^* + \delta_2^* - \delta_1^*, \theta_k = \theta_0 + \delta_2^*$  [2]. If we choose  $\delta_0^* = 0$  and our  $\delta_1 = \delta_2^* - \delta_1^*$ ,  $P(\Pi_k|LFC_1)$  is equivalent to our Power(LFC). Similarly, under these  $\delta$  values  $P(\Pi_0|LFC_0)$  is equivalent to  $1 - \alpha$  for our design.

We will define a few more values to make comparison easier. In these two selection procedures,  $n$  values are sampled from each population.  $c$  is similar to our  $y_1$ ; the number of successes from each population is compared to this number. In the one-stage, fixed sample size procedure,  $n$  observations are collected from all populations ( $N_T = nk$  total) and their numbers of successes are compared to each other and  $c$ . In the selection with curtailment procedure, one observation is taken from each population, then the researcher checks whether or not any of the populations can be selected or eliminated from consideration. If none can be eliminated, another observation is taken from each population. The expected sample size  $E_C[N]$  in the table below is for the curtailment procedure, which allows for the possibility of early termination before reaching  $N_T$ . It is found by averaging the expected sample size for the curtailment procedure under  $LFC_0$  and  $LFC_1$  [2]. We also denote the maximum sample size of our own procedure as  $N_{max} = kn_1 + n_2$ .



Table 3: Comparison of designs, for  $\theta_0 = 0.2$ ,  $\delta_0^* = 0$ ,  $\delta_1 = \delta_2^* - \delta_1^* = 0.05$ ,  $\delta_2 = \delta_2^* = 0.25$ , and  $\alpha = 1 - P_0^* = 0.1$ .

$k$	Power	Pure selection designs				Hybrid design					
		$n$	$c$	$N_T$	$E_C[N]$	$n_1$	$n_2$	$y_1$	$y_2$	$N_{max}$	$E[N]$
2	0.75	21	7	42	33.69	13	9	3	7	35	32.2425
	0.8	24	8	48	38.29	16	10	4	8	42	38.5452
	0.85	28	9	56	45.39	18	15	4	10	51	46.9915
	0.9	33	10	66	54.53	24	13	5	11	61	58.1688
	0.95	43	13	86	71.84	35	15	9	14	85	79.4551
3	0.75	26	9	78	31.86	16	12	3	9	60	58.6884
	0.8	27	9	81	65.12	18	18	4	11	72	68.5921
	0.85	31	10	93	76.05	24	15	6	12	87	82.8976
	0.9	38	12	114	94.61	30	17	7	14	107	103.2298
	0.95	49	15	147	123.93	43	12	10	16	141	138.2451
4	0.75	28	10	112	88.25	18	22	5	12	94	87.3447
	0.8	30	10	120	97.11	22	21	6	13	109	102.8125
	0.85	36	12	144	117.64	28	14	7	13	126	122.8258
	0.9	41	13	164	136.76	36	14	10	15	158	153.1065
	0.95	53	16	212	180.31	49	12	13	18	208	203.9930

Tables with the same comparison for different values of  $\theta_0$  are contained in the appendix. A similar pattern to the one we describe in the following two paragraphs can also be observed in those tables.

First we compare the fixed sample size procedure with our own. Because this procedure is single-stage, the experiment could be completed in half of the time it would take to perform the complete version of our procedure. So, if time is an issue, it may be more convenient to use the single-stage procedure. However, our two-stage design consistently has a lower overall sample size than the pure selection procedure. If the researchers have sample size constraints, it may make more sense to use our procedure.

Now we consider the curtailment procedure. While the same maximum sample size considerations apply to this design, the curtailment procedure also includes the possibility of early termination. Across most of the designs in the table, the expected sample size for the curtailment procedure is similar to or lower than our two-stage design, which is a definite advantage. On the other hand, the sampling method for curtailment forces researchers to only record results for one member of each population at a time. For example, in a  $k = 3$  situation, one response would be recorded from each of the three populations, at which point

the researcher would check to see if the experiment should be terminated, and then continue on to recording the next 3 responses. So, if the response takes a week to be recorded, our design would be completely finished in one to two weeks, while the curtailment procedure could continue for months. In the end, the choice between any of these procedures depends on the constraints under which researchers are operating.

### 3.3 Example

We will now give a brief example of using this design. The situation is as follows: researchers conducting clinical trials are interested in knowing whether or not any of three new pain medications have a success rate that is significantly more than 60%, with power 0.9. We assume they are operating under the same design restrictions that we have set, with  $\alpha = 0.05$ ,  $\delta_1 = 0.05$ , and  $\delta_2 = 0.2$ . In words, this means that they are only interested in a pain medication with an “acceptable” success rate of at least 80%, and not interested in medications with an “unacceptable” success rate of 65% or less. We pick the appropriate design from our table.

$k$	$\theta_0$	Power	$n_1$	$n_2$	$y_1$	$y_2$	Power	Size	$E[N]$	$\tau_0$
3	0.6	0.9	49	35	33	59	0.9000	0.0465	169.6553	0.6934

The researchers count a “success” as one of the subjects experiencing a large reduction in pain within one hour of taking the medication, and a “failure” as anything else. They assign 49 subjects to take each of the medications and the following results are recorded:

*Table 4:* Example experiment results

Treatment	Successes
1	41
2	30
3	35

The researchers select the first medication, because its number of successes is the highest, and is greater than  $y_1$ . They assign another 35 random people to the first medication and

record an additional 27 successes, for a total of 68 successes. Since  $y_2$  is 59, they conclude that the first medication is a significant improvement over the standard 60% success rate.

## 4 Conclusion

We have outlined a design for comparing binomial success rates between several populations in comparison with a standard. To reduce the overall sample size, the two-stage design allows for early termination of the experiment at the first stage if none of the experimental treatment seems promising. Power of the procedure is calculated under a least favorable configuration. This design could be used in clinical trials for comparing the effectiveness of several clinical treatments, or in any similar situation involving several experimental treatments with binary responses.

Because this is a hybrid, two-stage design, there are a few considerations that go into choosing to employ it. First, the two stages can restrict the type of experiment or observational study for which this would be practical. If it would take a year to record the responses in one stage, running two stages could take too much time. In such a case, it may be preferable to employ a single-stage design instead. However, if the researcher prefers to use a two-stage design, this one does have the advantage of allowing for early termination if none of the treatments appear promising. The two stages also allow researchers to not expend too many resources on sampling from populations that do not have sufficiently high success rates, instead only using the second stage to investigate the most promising population.

In the future, it may be interesting to extend our design to the possibility of multiple populations with a higher success rate, rather than just the one. Specifically, we could investigate the possibility of multiple populations getting through to the second stage in the event of a tie, rather than randomly selecting one of the populations. There is also the possibility of using preexisting knowledge about the treatments to break a tie. For example, if researchers were performing their experiment on human subjects and two of the treatments tie, it may be more reasonable to select the treatment with less severe side effects to continue

to the second stage.

It could also be worthwhile to study the behavior of this design with sample size restrictions, as mentioned in our comparison with Thall et. al's design [3]. Since we used the exact binomial distribution in our calculations, rather than a normal approximation, it would be possible to derive power, size, and early termination probabilities for very small sample sizes.

Rather than selecting the design with the lowest  $E[N]$ , it could be interesting to see how the design parameters compare when the designs are selected based on a different criterion. Our  $E[N]$  is defined as an average between  $E[N|H_0]$  and  $E[N|LFC]$ ; we could instead examine the behavior of  $E[N|\theta]$  under different values of  $\theta$ . Depending on the results of such an investigation, it may make sense to define  $E[N]$  under a single configuration. Alternatively, we could depart from using  $E[N]$  as the design selection criterion and instead select the design with the minimal total sample size (including the second stage). Finally, we could consider any one of these criteria, and also require the early termination probability  $\tau_0$  to be over a certain minimum amount.

Another way to extend this design could be the addition of a curtailment procedure, much like in [2]. If a certain number of successes or failures are observed in any of the populations, researchers would be able to save resources and stop sampling, especially in the first stage. However, this possibility would most likely only be feasible in experiments with quickly observable responses, since a two-stage design with curtailment would otherwise take a very long time to be completed.

# Appendix

## Comparison tables

*Table 5:* Comparison of designs, for  $\theta_0 = 0.4$ ,  $\delta_0^* = 0$ ,  $\delta_1 = \delta_2^* - \delta_1^* = 0.05$ ,  $\delta_2 = \delta_2^* = 0.25$ , and  $\alpha = 1 - P_0^* = 0.1$ .

$k$	Power	Pure selection designs				Hybrid design					
		$n$	$c$	$N_T$	$E_C[N]$	$n_1$	$n_2$	$y_1$	$y_2$	$N_{max}$	$E[N]$
2	0.75	25	14	50	36.86	13	16	6	15	42	36.5758
	0.8	28	15	56	43.01	17	16	8	17	50	44.3396
	0.85	30	16	60	46.41	19	18	8	19	56	51.8360
	0.9	38	20	76	30.09	27	14	11	21	68	65.3476
	0.95	47	24	94	76.08	40	12	18	26	92	88.2197
3	0.75	27	15	81	59.64	16	20	6	19	68	66.5043
	0.8	31	17	93	69.95	21	17	9	20	80	77.1204
	0.85	37	20	111	85.40	28	14	13	22	98	94.1411
	0.9	45	24	135	106.02	34	19	16	27	121	115.1686
	0.95	56	29	160	135.68	48	16	23	32	160	154.1884
4	0.75	32	18	128	93.52	20	24	9	23	104	66.5043
	0.8	36	20	144	107.32	25	21	12	24	121	77.1204
	0.85	42	23	168	128.15	31	23	15	28	147	64.1411
	0.9	47	25	188	147.61	40	16	19	29	176	115.1686
	0.95	60	31	240	194.35	54	15	25	35	231	154.1884

Table 6: Comparison of designs, for  $\theta_0 = 0.6$ ,  $\delta_0^* = 0$ ,  $\delta_1 = \delta_2^* - \delta_1^* = 0.05$ ,  $\delta_2 = \delta_2^* = 0.25$ , and  $\alpha = 1 - P_0^* = 0.1$ .

$k$	Power	Pure selection designs				Hybrid design					
		$n$	$c$	$N_T$	$E_C[N]$	$n_1$	$n_2$	$y_1$	$y_2$	$N_{max}$	$E[N]$
2	0.75	20	15	40	25.93	12	8	8	15	32	29.3586
	0.8	20	15	40	25.88	13	10	8	17	36	33.8219
	0.85	27	20	54	36.73	14	15	8	21	43	40.9863
	0.9	30	22	60	42.02	21	11	14	23	53	49.3805
	0.95	36	26	72	52.22	27	17	17	31	71	66.9211
3	0.75	25	19	75	46.96	13	14	8	20	53	51.0448
	0.8	25	19	75	46.96	15	21	10	26	66	60.6927
	0.85	28	21	84	54.89	21	13	14	25	76	72.5978
	0.9	34	25	102	70.14	27	10	18	27	91	88.2543
	0.95	44	32	132	94.10	36	13	24	35	121	117.1621
4	0.75	26	20	104	63.11	16	18	11	25	82	77.4022
	0.8	29	22	116	72.56	19	18	13	27	94	89.4131
	0.85	32	24	128	83.12	24	14	16	28	110	106.9787
	0.9	39	29	156	104.19	30	16	21	33	136	130.5065
	0.95	48	35	192	135.14	40	16	27	40	176	171.3742

## Model code

Note that this code will output a single design and must be adjusted depending on what sort of design is desired. The same is true of the simulation code.

---

*#dbinom(x,n,p) (pdf) or pbinom (cdf)*

*#packages needed: used "beepr" to play notification sound*

```
power_calc <- function(k,pn1,pn2,py1,py2,theta0,delta1,delta2){
  pow <- 0
  for (x in (py1+1:pn1)){ #power summation starting where the indicator term is
    1/nonzero
  }
  probm <- 0
  for(i in (0:(k-1))){ #probability of M event for power summation
    probm <- (probm + (1/(i+1))*choose(k-1,i)
      *((dbinom(x,pn1,theta0+delta1))^i)
      *((pbinom(x-1,pn1,theta0+delta1)))^(k-1-i))
  }
}
```

```

    }

    pow <- (pow + dbinom(x,pn1,theta0+delta2)
            *(1-pbinom(py2-x,pn2,theta0+delta2))
            *probm)

  }
  return(pow)
}

#vector of design parameters
design_param <- c(0,0,0,0,0,0)

ny <- c(0,0,0,0)
design_param[1] <- 2 #k
design_param[2] <- 0.6 #theta0
design_param[3] <- 0.9 #power
design_param[4] <- 0.05 #size
design_param[5] <- 0.05 #delta1
design_param[6] <- 0.2 #delta2

#example: design = (k=4, theta0=0.5, power=0.8, size/alpha = 0.05, delta1 =
0.05, delta2 = 0.2)

#ny = (n1=20, n2=15, y1=14, y2=8)

#initializing
power <- 0
size <- 1
test_design <- c(0,0,0,0,0,0,0,0)

start_time <- Sys.time()

```



```

for (n1 in (10:30)){ #looping through n1

for (n2 in (10:30)){ #looping through n2

for (y1 in ((round(n1*design_param[2])-1):n1)){ #looping through y1=n1*theta0
up to n1

for (y2 in (y1:(n1+n2))){ #looping through y2 = (n1+n2)*theta0 up to n1+n2
power <-
  power_calc(design_param[1],n1,n2,y1,y2,design_param[2],design_param[5],
    design_param[6])

if(power >= design_param[3] & power <= 1){ #if minimum is reached, size
triggers
  #print("power")
  #print(power)
  size <- 0 #size is k*power(delta1=0,delta2=0)
  size <- design_param[1] *
    power_calc(design_param[1],n1,n2,y1,y2,design_param[2],0,0)
  #print("size")
  #print(size)
}else{power <- 0} #reset value of power and start over **still
necessary after change?
#ny <- c(n1,n2,y1,y2)

if(size <= design_param[4] & size > 0){
  expected_n <- 0.5 * (2 * design_param[1] * n1 + n2 * (2 -
    (pbinom(y1,n1,design_param[2]))^design_param[1]
    -(pbinom(y1,n1,design_param[2] + design_param[5]))

```

```

      ^ (design_param[1]-1)
      * pbinom(y1,n1,design_param[2] + design_param[6]))
tau0 <- (pbinom(y1,n1,design_param[2]))^design_param[1] #probability of
      early termination
test_design <-
      rbind(test_design,c(n1,n2,y1,y2,power,size,expected_n,tau0))
size <- 1
power <- 0}else{
size <- 1
power <- 0}
}
}
}
}

```

```
test_design <- test_design[-1,]
```

```
min_n <- min(test_design[,7])
```

```
test_design[test_design[,7] == min_n,]
```

```
end_time <- Sys.time()
```

```
end_time - start_time
```

```
beep::beep()
```

---

## Simulation code

---

```
k <- 2
```

```
n1 <- 29
```

```

n2 <- 19
y1 <- 22
y2 <- 39
theta0 <- 0.7
delta1 <- 0.05
delta2 <- 0.2
n <- 100000
times_correct <- 0 #number of times the simulation correctly chooses the 2nd
    population
times_incorrect <- 0 #number of times the simulation falsely chooses the 1st
    population
times_failed <- 0 #number of times nothing is significant

for (i in (1:n)){
  sample_11 <- rbinom(1,n1,theta0 + delta1)
  sample_12 <- rbinom(1,n1,theta0 + delta2)

  if(sample_12 > y1 & sample_11 < sample_12){ #only 2 is significant
    sample_22 <- rbinom(1,n2,theta0+delta2)
    if(sample_12 + sample_22 > y2){ #2 is significant in stage 2
      times_correct <- times_correct + 1
    }else{ #nothing significant
      times_failed <- times_failed + 1
    }
  }else if(sample_11 > y1 & sample_12 < sample_11){ #only 1 is significant
    sample_21 <- rbinom(1,n2,theta0+delta1)
    if(sample_11 + sample_21 > y2){ #1 is significant in stage 2
      times_incorrect <- times_incorrect + 1
    }else{ #nothing significant

```

```

        times_failed <- times_failed + 1
    }
} else if(sample_11 > y1 & sample_12 > y1 & sample_11 == sample_12){ #both
    equal and significant
    if(sample.int(2,1) == 2){ #sample from 2 is picked
        sample_22 <- rbinom(1,n2,theta0+delta2)
        if(sample_12 + sample_22 > y2){ #2 is significant in stage 2
            times_correct <- times_correct + 1
        } else{ #nothing is significant
            times_failed <- times_failed + 1
        }
    } else{ #sample from 1 is picked
        sample_21 <- rbinom(1,n2,theta0+delta1)
        if(sample_11 + sample_21 > y2){ #1 is significant in stage 2
            times_incorrect <- times_incorrect + 1
        } else{ #nothing is significant
            times_failed <- times_failed + 1
        }
    }
}
} else{ #super not significant
    times_failed <- times_failed + 1
}
}

n
times_correct
times_incorrect
times_failed

```

---

## References

- [1] Miklós Bóna. *A Walk Through Combinatorics*. 4th ed. World Scientific Publishing Co. Pte. Ltd., 2017.
- [2] Elena M. Buzaianu. “Selection among Bernoulli populations in comparison with a standard”. In: *Sequential Analysis* 38.2 (2019), pp. 184–198.
- [3] Peter F. Thall; Richard Simon; Susan S. Ellenberg. “Two-Stage Selection and Testing Designs for Comparative Clinical Trials”. In: *Biometrika* 75.2 (1988), pp. 303–310.